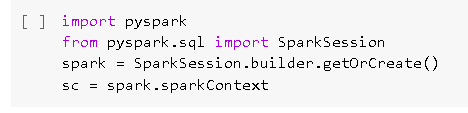
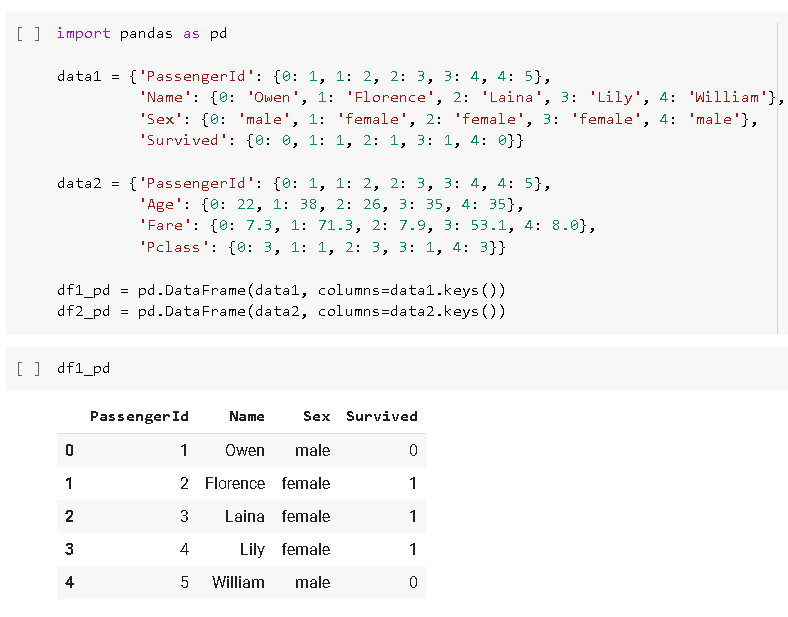
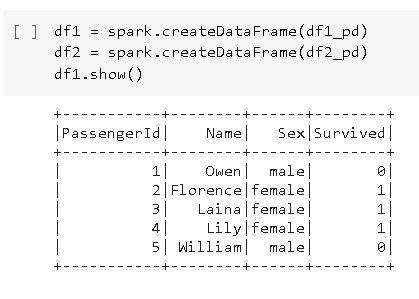
Làm việc với DataFrame Pyspark

Tạo môi trường pyspark:

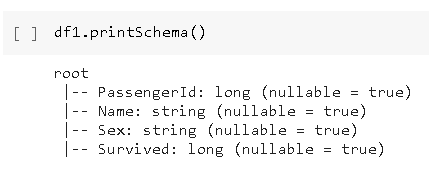


Convert dataframe từ pandas sang pyspark:



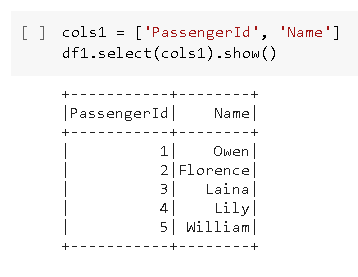


Xuất ra schema của dataframe:

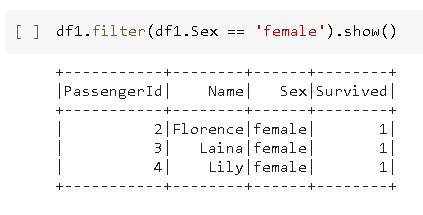


Các hành động cơ bản được thực thi với dataframe pyspark: select, filter, thay đổi hoặc tạo mới các cột (mutate, or creating new columns), tổng hợp và nhóm theo (summarize and group by), sắp xếp (sort), join và union,

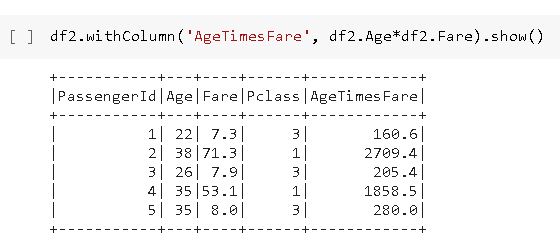
Select:



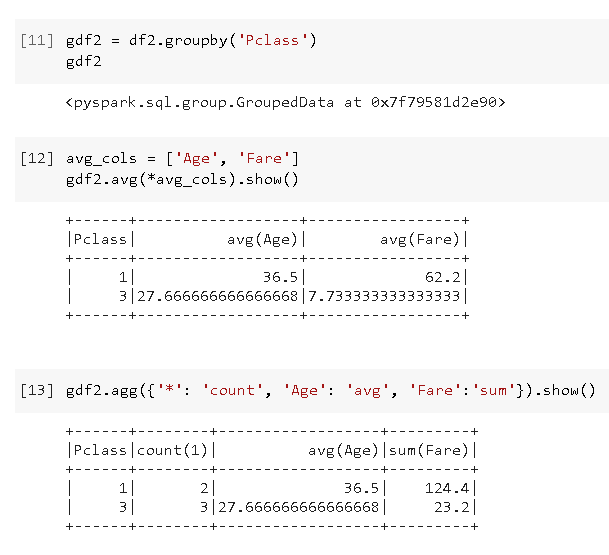
Filter:

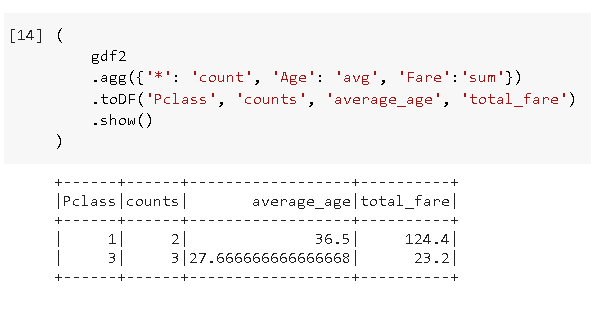


Tạo mới cột:

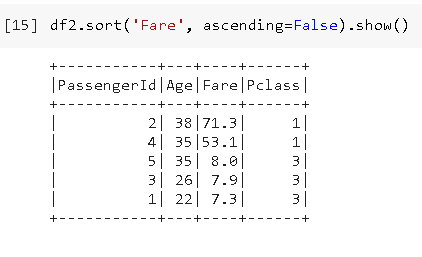


Group by:

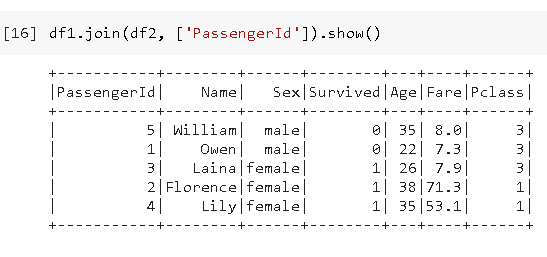




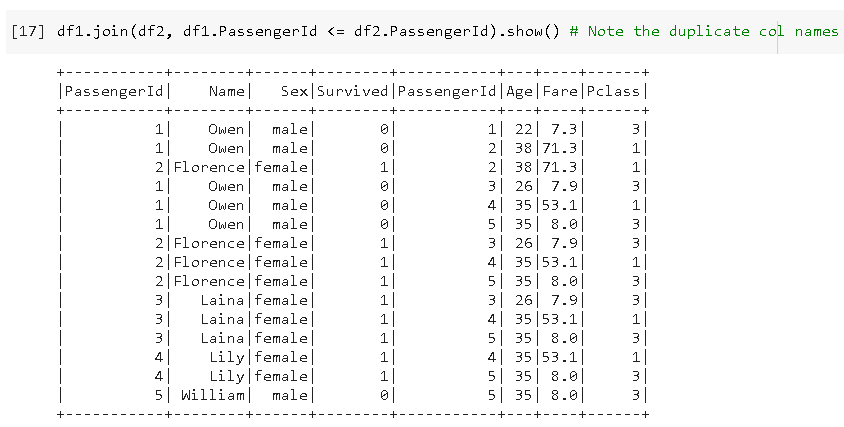
Sắp xếp:



Join:



Non-equi joins:



Union:

